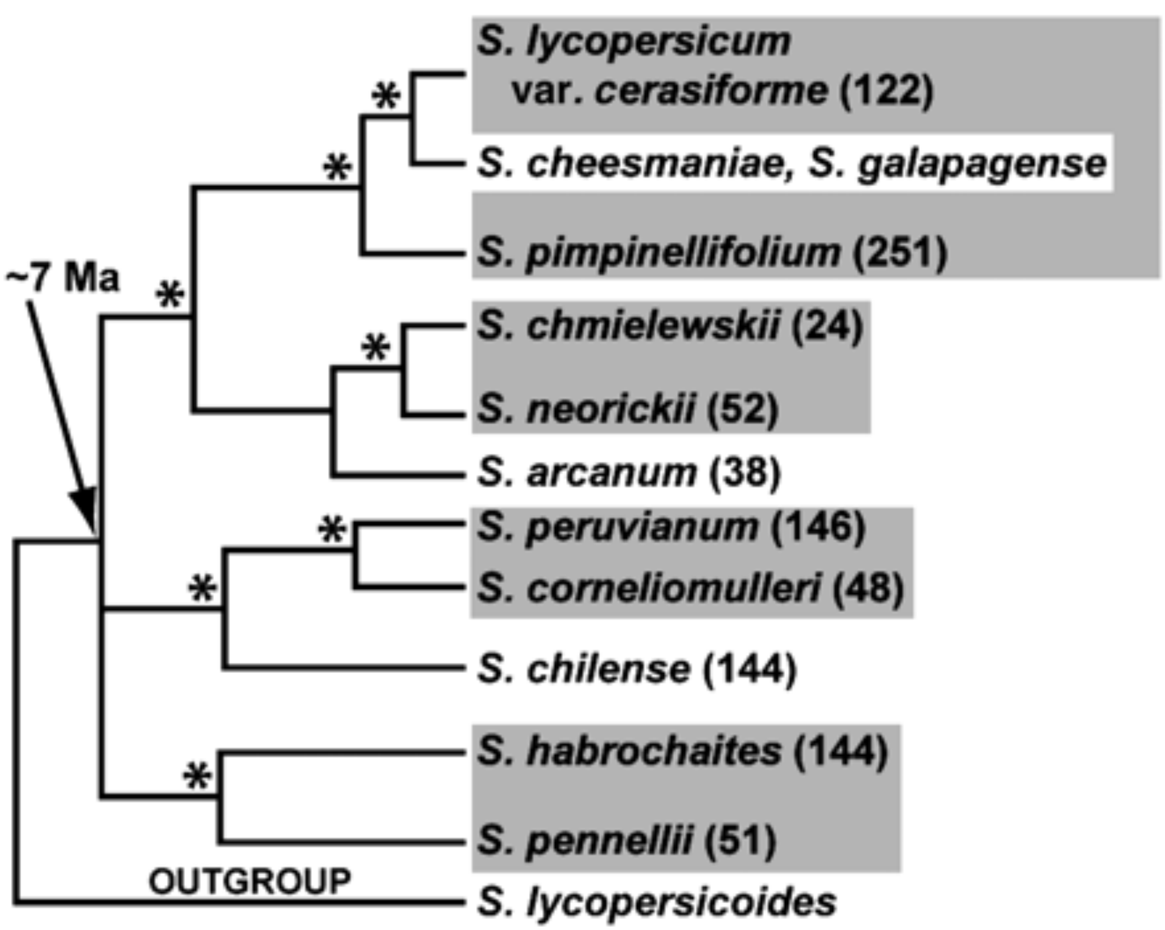
Allele Mining in Wild Tomato Species

Spooner et al., 2005

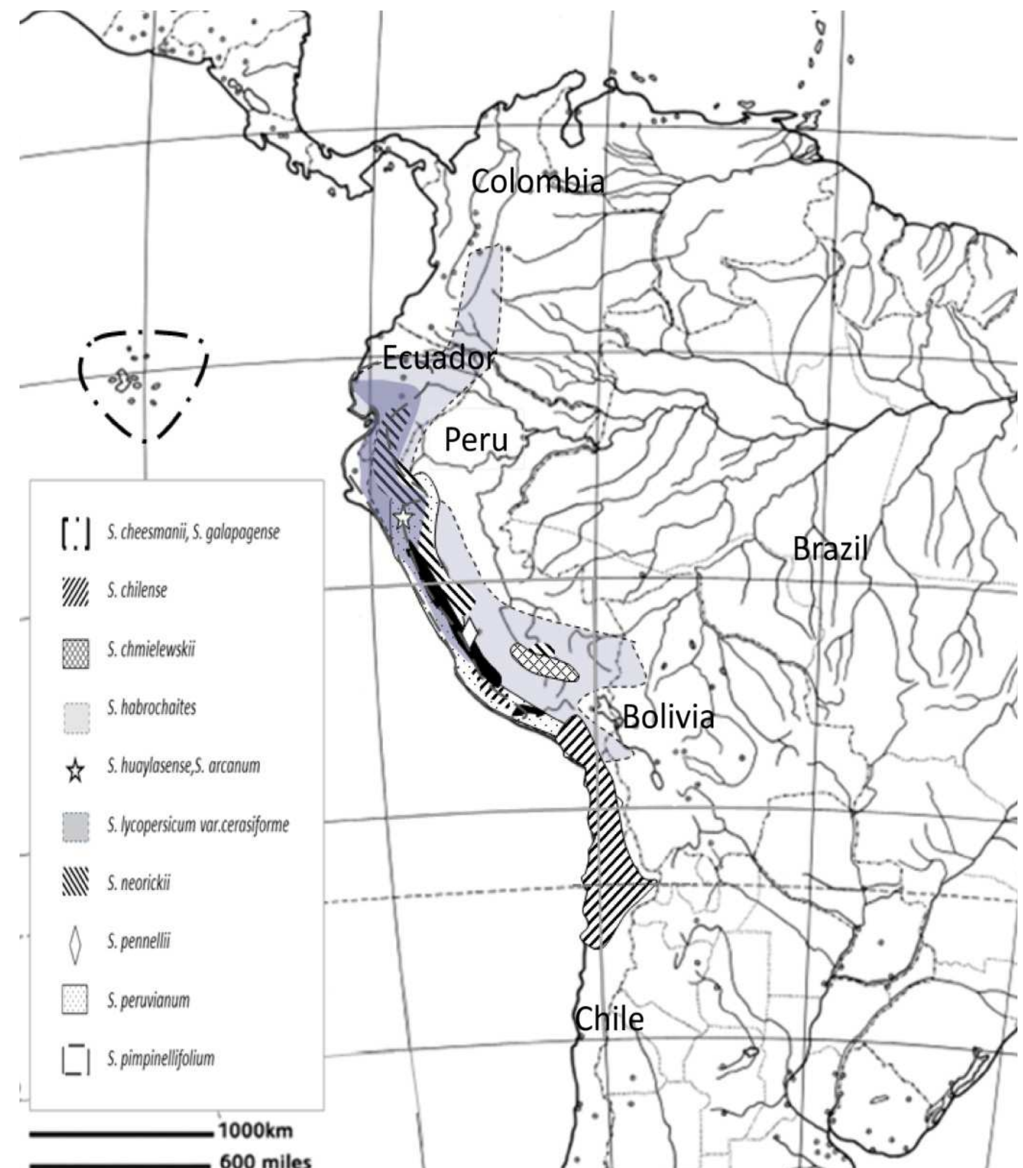
Susan Strickler
Mueller Lab

What are wild tomato species?

- 12 other species in Solanum section Lycopersicon
- Found in Western South America and Galapagos Islands



Nakazato et al., 2010



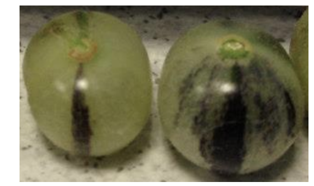
Bauchet and Causse, 2012

What are wild tomato species?

S. habrochaites

S. chilense

S. pimpinellifolium, *S. lycopersicum* 'Heinz', *S. habrochaites*



S. corneliomuelleri

S. cheesmaniae

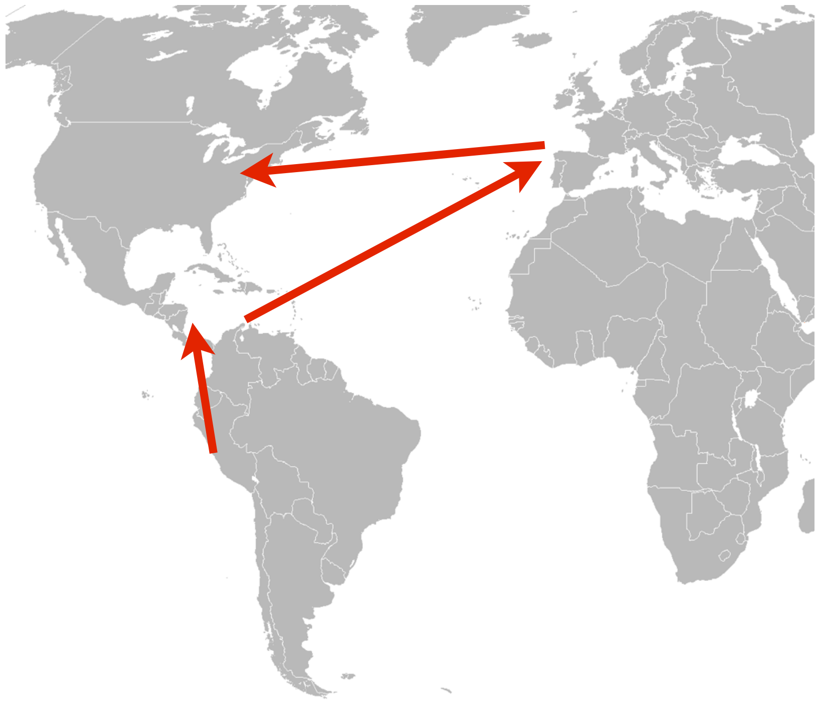
S. pimpinellifolium

S. chilense

Selection of fruit from TGRC.

Introgression in Nature and Breeding

- Crossing is feasible between cultivated and wild tomato and happens in nature (Grandillo et al., 2011).
- Much diversity present in wild tomato species. (Bai and Lindhout, 2007). Useful in breeding.
- Cultivated tomato has undergone three bottlenecks (Nesbitt and Tanksley, 2002).
- Differences in a few genes lead to many phenotypes.



Wild Tomato Species in Breeding

- Wild species have been used in breeding for:
  - disease resistance (ex: tobacco mosaic virus, tomato yellow leaf curl, bacterial canker, early blight, gray mold, late blight, leaf mold, powdery mildew)
  - fruit quality and nutritions (ex: brix, yield, flavor, ascorbic acid, fruit color, fruit shape, antioxidants)
  - other (ex, flowering time, leaf and flower morphology)

Gh13, early Oct. 2012

Douglas Maxwell

Project Goals

- Identify putative introgressions in the sequenced tomato from wild tomato species.
- Resolve Solanum phylogenies using genic and whole genome data
- Use approach to analyze introgression in breeding lines.

Whole Genome Sequence

- BTI-generated
  - *Solanum galapagense*
  - *Solanum lycopersicum* 'Yellow Pear'
- Publicly Available
  - *Solanum lycopersicum* 'Heinz'
  - *Solanum tuberosum*
  - *Solanum pimpinellifolium* (Lippman Lab)
- *Solanum lycopersicum* 'Heinz'

S. galapagense (LA0436)

Yellow Pear

S. pimpinellifolium (LA1589)

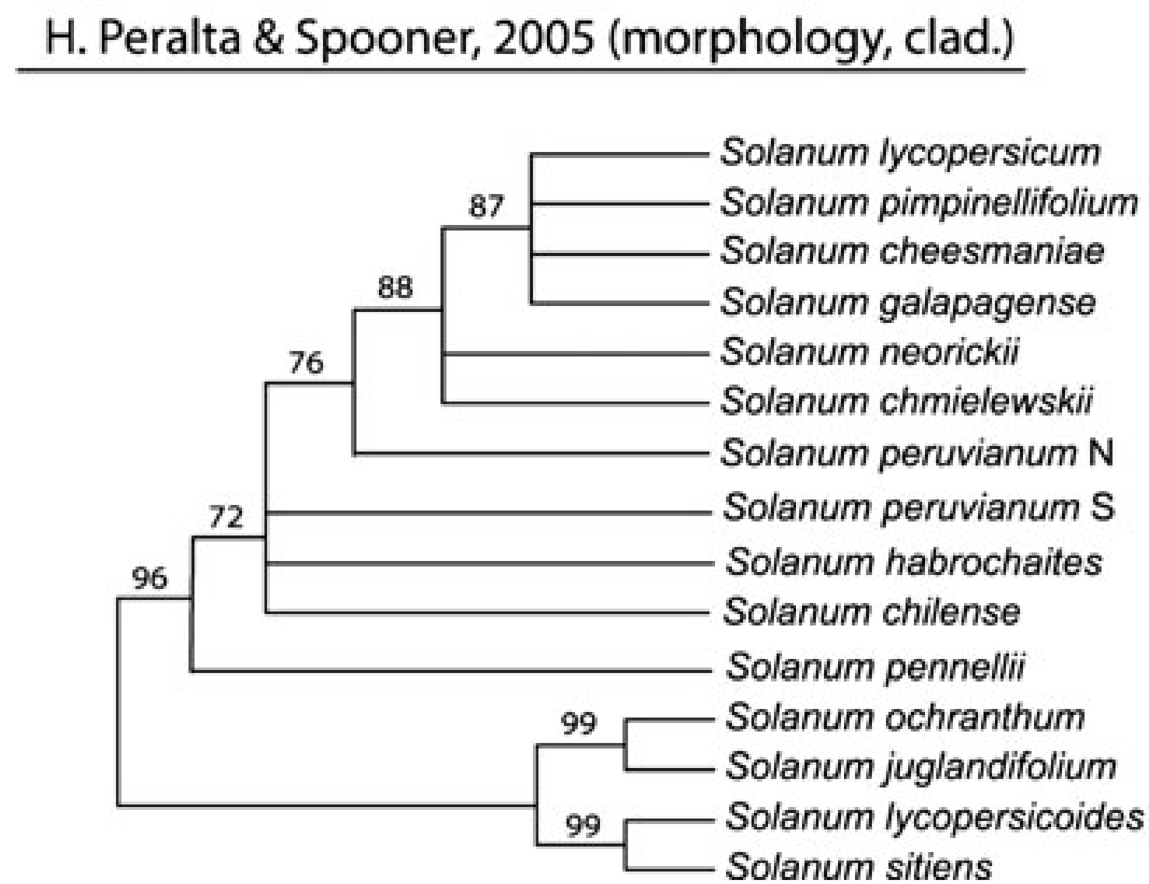
Why Sequence *S. galapagense* and Yellow Pear?

S. galapagense

- Has evolved in relative isolation from other tomato species until recently
- Closely related to *S. lycopersicum*, a good reference is available
- Unresolved position in *Solanum* phylogenetic tree - closely related to *S. cheesmaniae* and only recognized as a separate species recently.

Yellow Pear

- Old heirloom line of *S. lycopersicum*. Should be relatively introgression-free.



Silvana et al., 2011

Sequence Assembly Pipelines

Reference sequence:

Tomato Geome Consortium, 2012

Illumina Sequencing
LA436
(MP & PE)

align to Heinz
BWA & Novoalign

SNP/indel calling
GATK

SNP Effect
SNPeff

Whole Genome Aignment
Mercator & Mavid

De novo assembly
SOAP de novo

Align contigs to Heinz
Blat

sol genomics network

Reference-guided Assembly Statistics

	S. lycopersicum		Wild Species	
Reference-guided assembly	**Heinz**	**Yellow Pear**	***S. galapagense***	***S.pimpinellifolium***
filtered reads in millions	1,002.7	420.3	363.9	281.5
mapped reads in millions	834.1	393.1	324.7	247.7
coverage depth after dup and mq <30 removal	80x	45x	32x	25x
coverage of tomato gen[1]	99.3%	**99.3%**	**95.4%**	**95.0%**
no of gaps (total size in mb)[2]	60, 214 (5.6)	**51,980 (5.4)**	**227,699 (36.1)**	**209,919 (38.9)**
no of gaps > 500 bp*	**1,693**	**1,926**	**10,751**	**14,396**

[2]Gaps in Heinz assembly removed from calculation

Coverage of Heinz genome decreases

Sequence Assembly Pipelines

Reference sequence:

Tomato Genome Consortium, 2012

Illumina Sequencing
LA436
(MP & PE)

→ **align to Heinz**
BWA &
Novoalign

→ **SNP/indel calling**
GATK

→ **SNP Effect**
SNPeff

→ **Whole Genome Aignment**
Mercator &
Mavid

→ **De novo assembly**
SOAP de
novo

→ **Align contigs to Heinz**
Blat

Gap Analysis

	S. lycopersicum	Wild Species	
	Yellow Pear	***S. galapagense***	***S.pimpinellifolium***
unmapped contigs	11,841	15,067	22,346
Putative deleted regions (bp)*	951,969	13,812,215	21,599,274
Putative deleted genes	13	87	157

*Putative deleted region = gap in mapping assembly and in blat analysis

Heinz gaps removed from analysis

Coverage of Heinz genome decreases

S. pimpinellifolium has more deleted regions and missing genes.

Reference-guided Assembly
SNP Locations

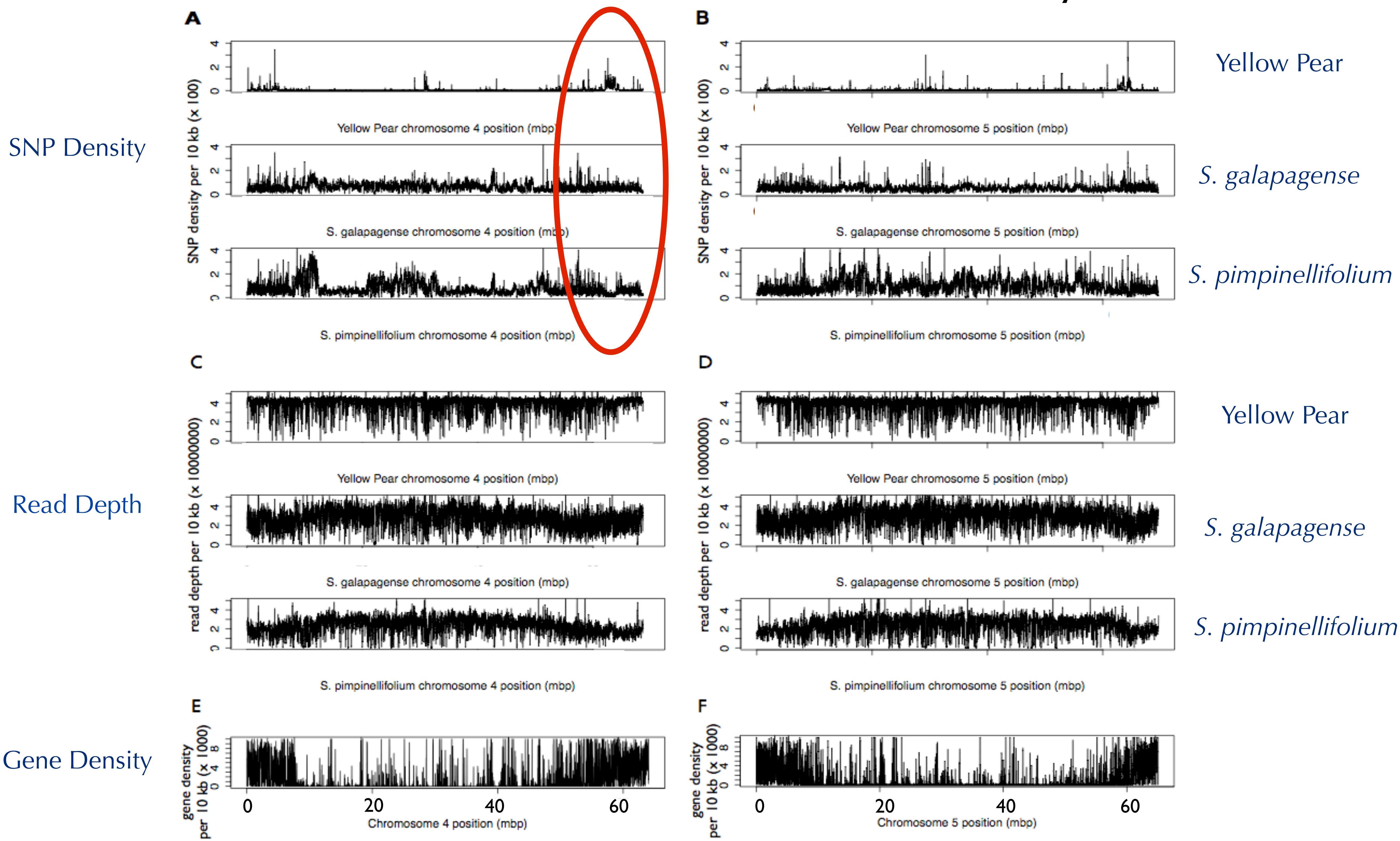


	# of SNPs (% of region)*		
SNP Location	**Yellow Pear**	***S. galapagense***	***S. pimpinellifolium***
Total SNPs	539,406 (0.07%)	4,665,765 (0.6%)	6,016,177 (0.8%)
Intergenic	339,858 (0.05%)	4,318,821 (0.5%)	5,589,364 (0.8%)
Genic	42,117 (0.04%)	346,944 (0.3%)	426,813 (0.4%)
Genic - noncoding (introns)	26,025 (0.04%)	241,551 (0.3%)	304,830 (0.4%)
Genic - coding	16,092 (0.04%)	105,393 (0.3%)	121,983 (0.3%)
coding nonsynonymous	10,118	62,432	70,474
coding synonymous	15,457	40,622	40,703

#of SNPs increases

S. galapagense has greater sequence similarity to Heinz than *S. pimpinellifolium*.

Reference-guided Assembly
Genome-wide SNP density

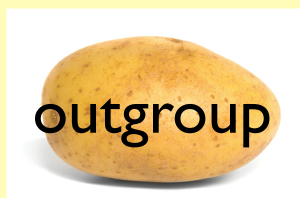


Phylogenomics

sol genomics network



Coding Sequence

Select coding sequence with > 50% coverage

Reciprocal BLAST to *S. lycopersicum* to determine orthologs

Is *S. galapagense* more closely related to *S. lycopersicum* than *S. pimpinellifolium?*

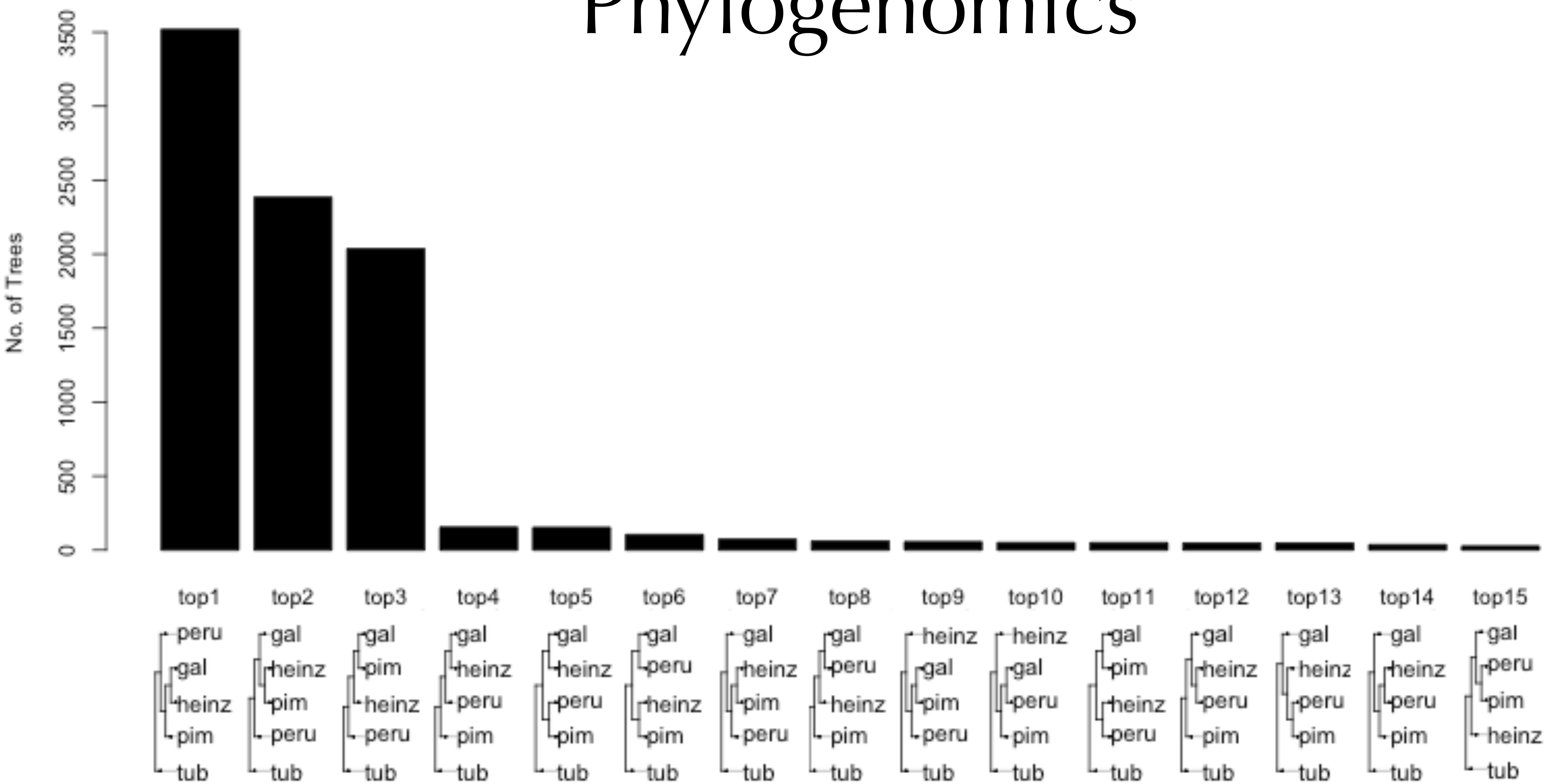
PhyGomics Pipeline
(Bombarely *et al.*, submitted)

ClustalW alignments
8,797

Maximum likelihood (phyml) used to infer phylogenetic relationships

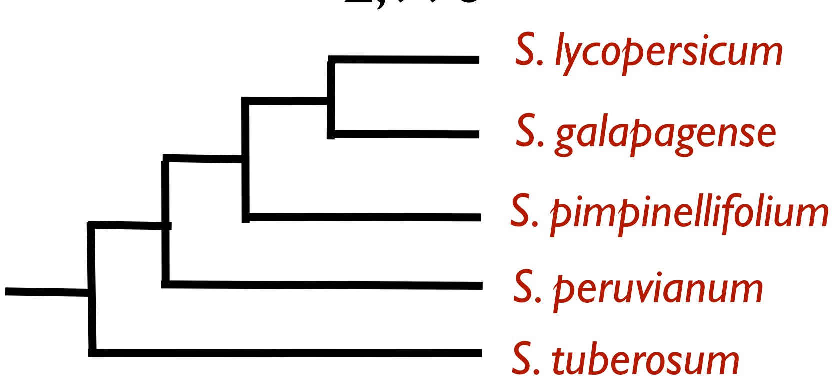
sol genomics network

Phylogenomics



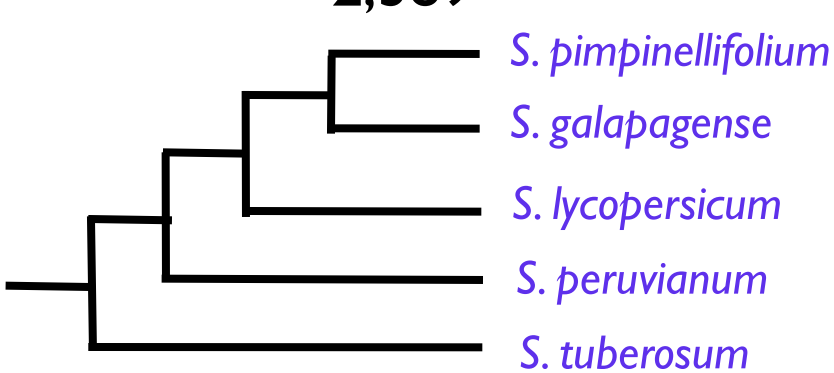
Topology 1

2,996



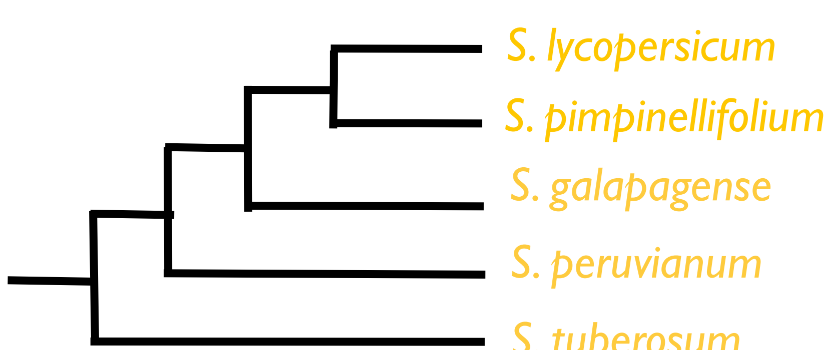
Topology 2

2,389



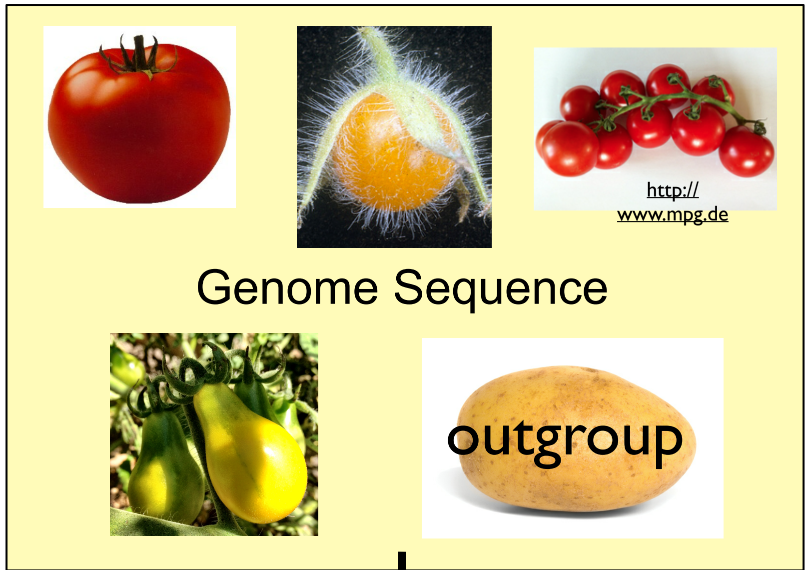
Topology 3

2,012



Whole Genome Phylogenies Pipeline

(adapted from White et al., 2009)



↓

Create genome using Heinz as template - substitute in gaps and SNPS.

↓

Mask repeats with Repeat Masker.

→

Whole genome alignments using Mercator and Mavid.

→

PAUP to deduce phylogenetic breakpoints

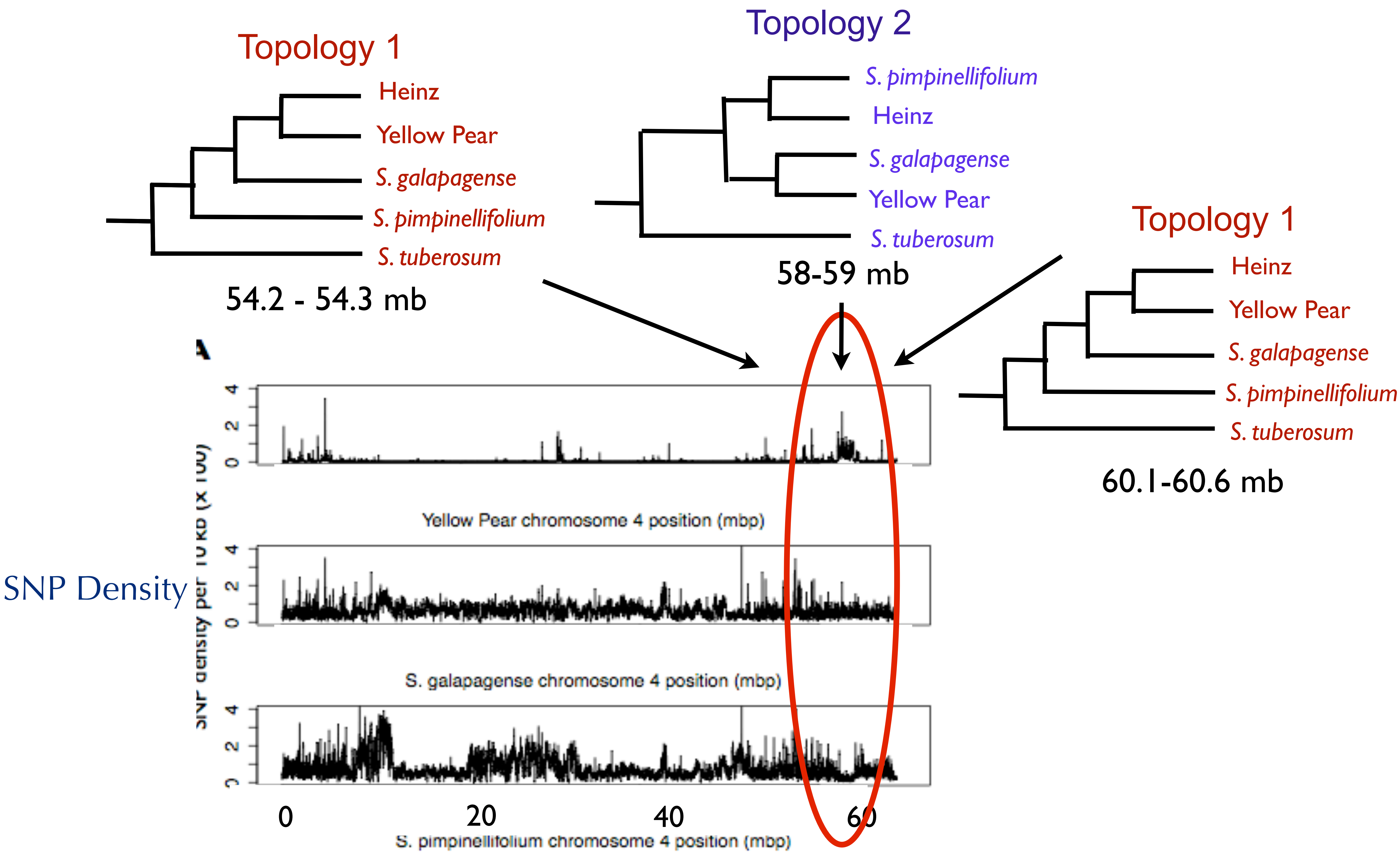
↓

Create trees for each partition with Mr. Bayes.

←

Map trees to Heinz genome.

Whole Genome Phylogenies



Whole genome sequencing of an inbred disease-resistant line.

- Inbred line with know introgression(s) for tomato yellow leaf curl (Ty) resistance
- *S. chilense*

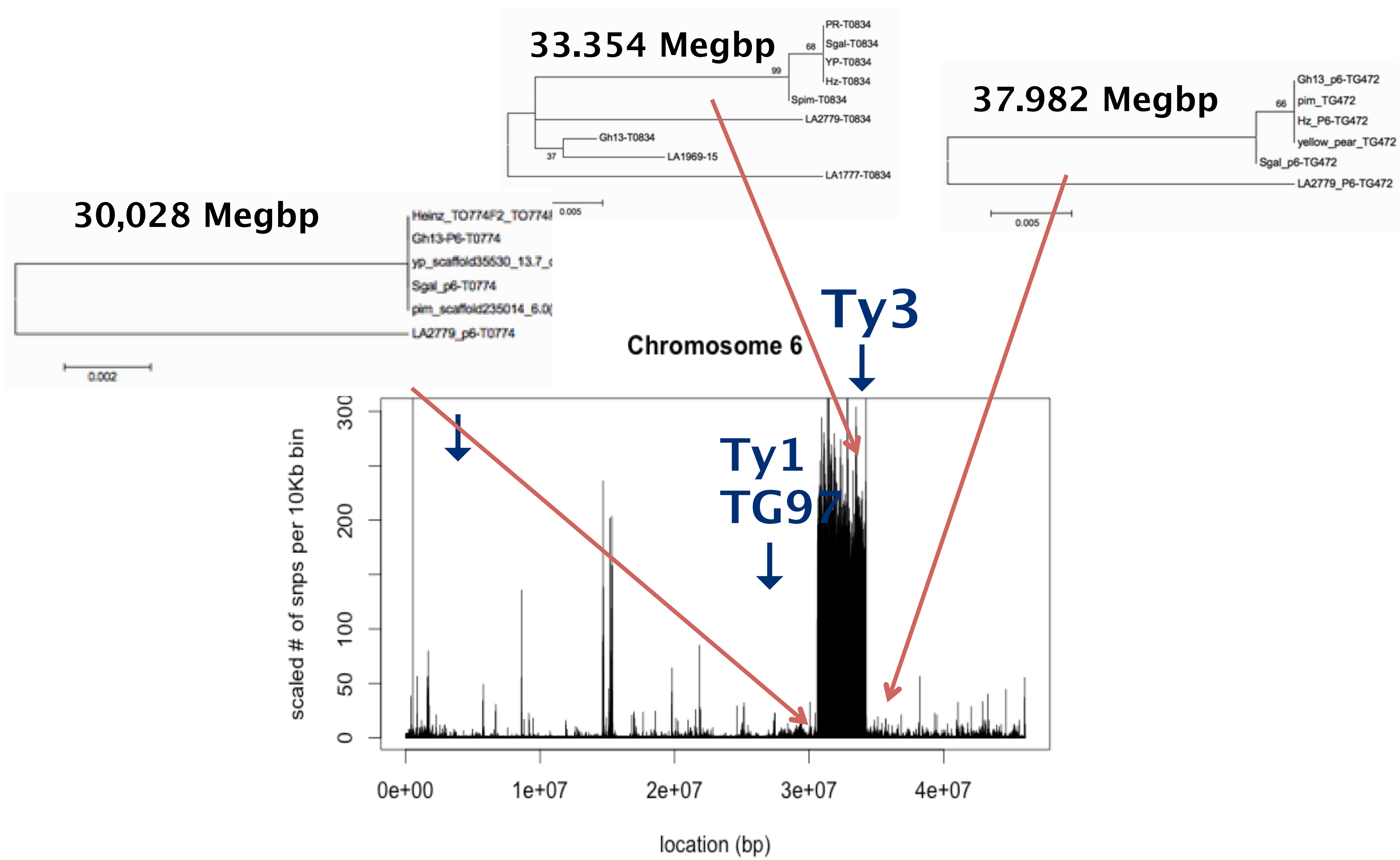
Gh13, early Oct. 2012

Commercial Hybrid

Naama Menda, Douglas Maxwell

sol genomics network

Inbred Results



Introgression = 30.6 – 34.213 Megbp

Naama Menda, Douglas Maxwell

Conclusions

- *S. galapagense* and *S. lycopersicum* have high sequence similarity, allowing *S. lycopersicum* tools to be implemented
- There is 99% sequence identity between *S. galapagense* and *S. lycopersicum*. SNPs and indels occur mainly in intergenic regions as expected.
- An introgression from *S. pimpinellifolium* may be present in *S. lycopersicum* 'Heinz' chromsome 4.
- Coding sequence analysis suggests *S. galapagense* may be the closest wild relative to *S. lycopersicum*. Current work involves constructing whole genome phylogenies to look at phylogenetic discordance and introgression.
- This approach is successful in identifying introgressions based on known introgression in inbred lines.

Acknowledgements

Mueller Lab

Aureliano Bombarely, Naama Menda

Greg Martin and Jesse Munkvold

Douglas Maxwell

Lippman Lab